PUBLISHED BY IOP PUBLISHING FOR SISSA
RECEIVED: *April 9, 2009*
ACCEPTED: *June 4, 2009*
PUBLISHED: *July 13, 2009*

4th INTERNATIONAL CONFERENCE ON IMAGING TECHNOLOGIES IN BIO MEDICAL SCIENCES,
FROM MEDICAL IMAGES TO CLINICAL INFORMATION - BRIDGING THE GAP,
22–28 SEPTEMBER 2007,
MILOS ISLAND, GREECE

2009 JINST 4 P07013

Exploiting unsupervised and supervised classification for segmentation of the pathological lung in CT

P. Korfiatis,[a] C. Kalogeropoulou,[b] D. Daoussis,[c] T. Petsas,[b] A. Adonopoulos[c] and L. Costaridou[a,1]

[a] *University of Patras, School of Medicine, Department of Medical Physics, Rio, Patras 265 00, Greece*
[b] *Department of Radiology, University Hospital of Patras, Rio, Patras 265 00, Greece*
[c] *Department of Rheumatology, University Hospital of Patras, Rio, Patras 265 00, Greece*

E-mail: costarid@upatras.gr

ABSTRACT: Delineation of lung fields in presence of diffuse lung diseases (DLPDs), such as interstitial pneumonias (IP), challenges segmentation algorithms. To deal with IP patterns affecting the lung border an automated image texture classification scheme is proposed. The proposed segmentation scheme is based on supervised texture classification between lung tissue (normal and abnormal) and surrounding tissue (pleura and thoracic wall) in the lung border region. This region is coarsely defined around an initial estimate of lung border, provided by means of Markov Radom Field modeling and morphological operations. Subsequently, a support vector machine classifier was trained to distinguish between the above two classes of tissue, using textural feature of gray scale and wavelet domains. 17 patients diagnosed with IP, secondary to connective tissue diseases were examined. Segmentation performance in terms of overlap was 0.924 ± 0.021, and for shape differentiation mean, rms and maximum distance were 1.663 ± 0.816, 2.334 ± 1.574 and 8.0515 ± 6.549 mm, respectively. An accurate, automated scheme is proposed for segmenting abnormal lung fields in HRC affected by IP

KEYWORDS: Data processing methods; Pattern recognition, cluster finding, calibration and fitting methods; Computerized Tomography (CT) and Computed Radiography (CR)

[1] Corresponding Author.

doi:10.1088/1748-0221/4/07/P07013

Contents

1 Introduction

A number of algorithms have been proposed for lung border delineation accuracy [1]–[5]. Most of these techniques have been tested on normal lungs with the exception of [3, 4], who dealt with focal pathologies attached to lung borders. Recently, the challenge of segmenting lung fields in presence pathologies affecting lung border is highlighted by Sluimer et al. [5].

Among lung pathologies, a special category is interstitial pneumonias, whose quantification and characterization is a difficult task for radiologists, due to complexity and variability of their morphologic patterns.

High resolution computed tomography (HRCT) is the choice to image Diffuse Parenchymal Lung Disease (DPLD) such as interstitial pneumonias (IP). Several computer aided diagnostic schemes for DLPDs on HRCT have been proposed [6]–[8]. These systems focus on the final stages of lung analysis, utilizing manual or automated thresholding techniques in a suboptimal way for lung field delineation. This may have serious consequences in the analysis steps that follow, due to exclusion of DLPDs regions or inclusion of surrounding lung tissue areas.

In this work, a segmentation algorithm is proposed for delineating abnormal lung border. The system exploits the combination of two increasingly complex segmentation methods. The Markov Random Field (MRF) unsupervised segmentation technique was chosen to provide an initial estimate of the lung border. This estimate in combination with morphological operations was utilized to define a lung border search region. The final segmentation result was obtained by exploiting texture differentiation between lung tissue (normal and abnormal) and lung surrounding tissue (pleura and thoracic wall) in this region. Two categories of textural features (1st and 2nd order statistics), were extracted from original gray level and wavelet coefficient domains. Following, a support vector machine classifier is used to distinguish lung tissue from surrounding tissue resulting in the final delineation of lung border. The proposed system is applied on a dataset containing HRCT images

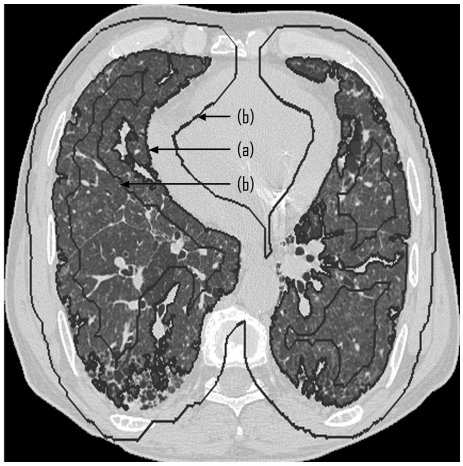


Figure 1. Arrows indicate: (a) Initial lung border produced by MRF segmentation, (b) inner and outer border of the LBR region.

with complex pathologies affecting lung borders. The performance of the proposed method is assessed using quantitative metrics, by comparing automatically derived lung borders to the manually traced borders from an experienced radiologist.

2 Methods and materials

2.1 Lung Border Segmentation (LBR)

The first step of the method utilizes an unsupervised segmentation scheme to coarsely define a lung border search region on which further segmentation using textural features and supervised classification is exploited for lung border refinement. In this region characteristic IP image patterns affecting segmentation accuracy are expected to be found.

An unsupervised MRF based segmentation algorithm was implemented to provide an initial estimate of lung border. Image segmentation using MRF provides the advantage of seamless integration of spatial relationships with various image features, into a segmentation procedure. The technique proposed by Deng and Clausi [9] is adopted using a function-based parameter to weight the two components (spatial relations and image features) in a MRF model to produce a more accurate segmentation result.

Morphological operations were applied on the initial estimate of the lung border, as identified MRF labelling to define the LBR. Specifically dilation using a circular structural element, whose radius was empirically selected to be 10 pixels was applied to define the outer border of the LBR, while erosion was applied to define it's inner border. Figure 1 demonstrates the defined LBR of a characteristic case of the sample analyzed.

2.2 Lung Border Refinement by local texture analysis

Following lung border region identification, a supervised classification scheme was implemented to distinguish between lung parenchyma (normal and abnormal) and surrounding tissue (fat muscle

2009 JINST 4 P07013

and bones) based on image features extracted from a sliding ROI in LBR region. A decision for the central pixel of the sliding ROI regarding tissue category is obtained by means of SVM.

From each ROI (9x9 pixels), features related to gray level and wavelet domain texture were extracted with the latter aiming to capture edge information. Specifically, 5 first order statistics were derived from the local histogram (mean, standard deviation, skewness, entropy and kurtosis). 13 second order statistics were computed from the local co-occurrence matrix [10] using one pixel step. Each second-order statistic feature was represented by two values, the mean and range over the 0°, 45°, 90° and 135° the co-occurrence (13×2 features). In total 62 textural features were calculated for each ROI.

A non linear Support Vector Machine classifier was used with a Gaussian kernel. All the parameters for the SVM classifier were automatically derived according to Radius/Margin bound [11]. The SVM was chosen as in case of non-linearly separable classes, the SVM classifier first maps the patterns to a higher-dimension feature space using a kernel transformation. A training sample of 300 ROIs extracted from LBRs, corresponding to 5 scans not included in the dataset, were used to train the classifier.

An effort was made to reduce feature dimensionality, aiming to potentially increase class separation and reduce time need for SVM training, by means of feature ranking. Features were ranked using Student's t-test. The top 10 were selected and exhaustive search was carried out. The classifier was designed using every possible feature-combination, among the top 10 selected features, each time testing the classifier's accuracy with the leave one out method. The feature-combination that demonstrated the highest classification accuracy with the smallest number of features was selected. These features were: mean of sum average, mean of sum variance, mean of difference entropy, mean gray level and skewness calculated from original image and mean of sum entropy calculated on wavelet magnitude coefficient A training accuracy of Az=0.9986$\pm$0.0083was obtained.

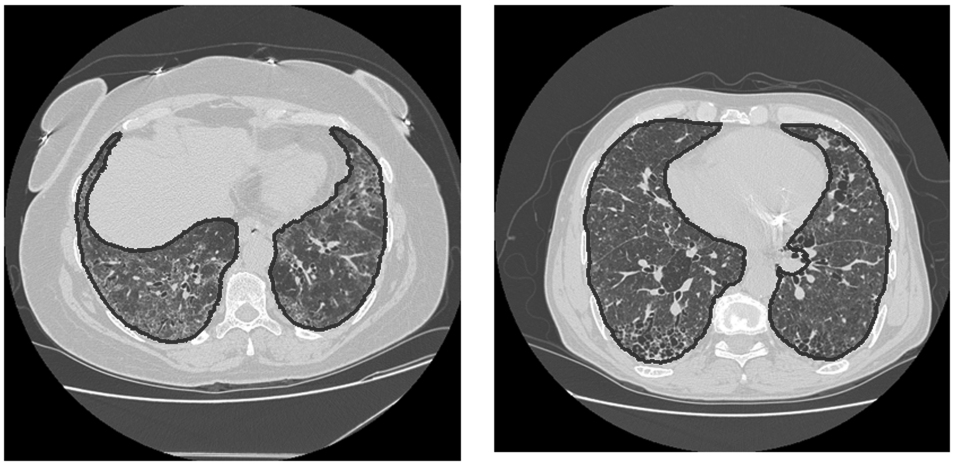
From each location of the sliding ROI in the LBR the above 5 features were extracted and used as input to the SVM classifier along with the training sample. The output of the classifier was a decision regarding the central pixel of the sliding block thus providing the final lung border. Figure 2 provides representative examples of lung border segmentation with the proposed method for (figure 2a) a honeycombing pattern and (figure 2b) a combined pattern of groundglass opacities, interstitial thickening, traction bronhiectases and honey combing.

2.3 Dataset

17 patients diagnosed with IP, secondary to connective tissue diseases were examined using an HRCT protocol (LightSpeed16 GE) in the department of Radiology at the University Hospital of Patras. 120 slices depicting abnormal lung boundary were selected for analysis. The image matrix size was 512×512 and mean pixel size 0.6 mm.

3 Results and discussion

In this work, an automated method for lung field segmentation was developed and validated. The method combines an unsupervised MRF technique to provide an initial estimate of the lung border. The initial lung border along with morphological processing is used to define a search border region (LBR), mostly affected by IP patterns. A supervised texture analysis scheme, based on a



(a) (b)

Figure 2. Two examples of lung border segmentation (a) Honeycombing (b) a case with Groundglass opacities, Interstitial thickening, Traction bronhiectases and findings of Honeycombing.

SVM classifier is applied on the LBR, and used to distinguish lung tissue from surrounding tissue resulting in the final delineation of lung border.

The accuracy of segmentation algorithm is evaluated using quantitative metrics, by comparing automatically derived lung borders to the manually traced borders from an experienced radiologist. Segmentation performance in terms of overlap was 0.924 ± 0.021, and for shape differentiation mean, rms and maximum distance were 1.663 ± 0.816, 2.334 ± 1.574 and 8.0515 ± 6.549 mm, respectively.

The processing time required to segment a slice, using a 2.6 GHz core duo Intel processor with 2 GB RAM, is approximately 8 min.

Comparison of method performance with other recently proposed lung field segmentation methods for CT data cannot be directly achieved owing to different datasets utilized. A recently reported atlas based lung segmentation scheme [5], tested on a pathological database containing 10 characteristic cases, reported volume overlap and mean absolute surface distance of 0.93 ± 0.04 and 1.48 ± 1.22 mm respectively.

Finally, improvement of the proposed method should deal with a more robust automatic identification of the LBR, possibly taking into account anatomical landmarks, as well as experimentation with additional textural features.

4 Conclusions

An automated lung segmentation algorithm was developed combining unsupervised and supervised techniques. The proposed method was tested on a data set containing HRCT images of patients diagnosed with interstitial pneumonia. Results demonstrate an accurate system both in terms of volume and shape. The method could be used as an initial step in applications aimed at computerized classification and quantification of diffuse lung disease abnormalities.

2009 JINST 4 P07013

Acknowledgments

This work is supported by the Caratheodory Programme (C.180) of the University of Patras

References

[1] S. Hu, E.A. Hoffman and J.M. Reinhardt, *Toward automated segmentation of the pathological lung in CT*, *IEEE Trans. Med. Imag.* **20** (2001) 490.

[2] X. Sun, H. Zhang and H. Duan, *3D computerized segmentation of lung volume with computed tomography*, *Acad. Radiol.* **13** (2006) 670.

[3] S.G. Armato III and W.F Sensakovic, *Automated lung segmentation for thoracic CT: impact on computer-aided diagnosis*, *Acad. Radiol.* **11** (2004) 1011.

[4] P. Korfiatis, S. Skiadopoulos, P. Sakellaropoulos, C. Kalogeropoulou and L. Costaridou, *Automated 3D Segmentation of Lung Fields in Thin Slice CT Exploiting Wavelet Preprocessing*, *Lect. Notes Comput. Sci.* **S 4673** (2007) 237.

[5] P. Korfiatis, S. Skiadopoulos, P. Sakellaropoulos, C. Kalogeropoulou and L. Costaridou, *Combining 2D wavelet edge highlighting and 3D thresholding for lung segmentation in thin-slice CT*, *Br. J. Radiol.* **80** (2007) 996.

[6] I. Sluimer, M. Prokop and B. Van Ginneken, *Toward automated segmentation of the pathological lung in CT*, *IEEE Trans. Med. Imag.* **24** (2005) 1025.

[7] I. Sluimer, M. Prokop, I.Hartmann and B.Van GInneken, *Automated classification of hyperlucency, fibrosis, ground glass, solid, and focal lesions in high-resolution CT of the lung*, *Med. Phys.* **33** (2006) 2610.

[8] Y. Uchiyama et al, *Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography*, *Med. Phys.* **30** (2003) 2440.

[9] H. Deng and D.A. Clausi, *Unsupervised segmentation of synthetic aperture radar sea ice imagery using a novel Markov random field model*, *IEEE Trans. Geosci. Remote Sens.* **43** (2005) 528.

[10] M. Haralick, K. Shanmugam and I. Dinstein, *Textural features for image classification*, *IEEE Trans. Syst. Man Cybern.* **3** (1973) 610.

[11] O. Chapelle, V. Vapnik, O. Bousquet and S. Mukherjee, *Choosing Multiple Parameters for Support Vector Machines*, *Mach. Learn.* **46** (2002) 131.

2009 JINST 4 P07013